Our results confirm that UL97 is an unusual member of the protein kinase sequence family. Characterization of the HCMV protein responsible for phosphorylating ganciclovir will help in the design of improved antiviral nucleoside analogues for treating HCMV infection. □

Received 11 February; accepted 27 May 1992.

1. Macher, A. M. *et al. New Eng. J. Med.* **309**, 1454 (1983).
2. Biron, K. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2473-2477 (1985).
3. Biron, K. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 7769-8773 (1986).
4. Freitas, V. R., Smee, D. F., Chernow, M., Boehme, R. & Matthews, T. R. *Antimicrob. Ag. Chemother.* **28**, 240-245 (1985).
5. Estes, J. E. & Huang, E. S. *J. Virol.* **24**, 13-21 (1976).
6. Fyfe, J. A., Keller, P. M., Furman, P. A., Miller, R. L. & Elion, G. B. L. *J. biol. Chem.* **253**, 8721-8727 (1978).
7. Chee, M. S. *et al. Curr. Top. Microbiol. Immun.* **154**, 125-169 (1990).
8. Chee, M. S., Lawrence, G. L. & Barrell, B. G. *J. gen. Virol.* **70**, 1151-1160 (1989).
9. Edelman, A. M., Blumenthal, D. K. & Krebs, E. G. *A. Rev. Biochem.* **56**, 567-613 (1987).
10. Pearson, R. B., Wettenhall, R. E. H., Means, A. R., Hartshorne, D. J. & Kemp, B. E. *Science* **241**, 970-973 (1988).
11. Rosenberg, A. H. *et al. Gene* **56**, 31-38 (1987).
12. Ertl, P. F., Thomas, M. S. & Powell, K. L. *J. gen. Virol.* **72**, 1729-1734 (1991).
13. Sullivan, V. *et al. Nature* **358**, 162-164 (1992).
14. Stanat, S. C. *et al. Antimicrob. Ag. Chemother* **35**, 2191-2197 (1991).
15. Smith, R. F. & Smith, T. F. *J. Virol.* **63**, 450-455 (1989).
16. Knighton, D. R. *et al. Science* **253**, 414-420 (1991).
17. Littler, E., Lawrence, G., Liu, M.-Y., Barrell, B. G. & Arrand, J. R. *J. Virol.* **64**, 714-722 (1990).
18. Littler, E. & Arrand, J. R. *J. Virol.* **62**, 3892-3895 (1988).

ACKNOWLEDGEMENTS. We thank D. Morris for clinical sera and S. Stanat for help in preparing the manuscript.

A protein kinase homologue controls phosphorylation of ganciclovir in human cytomegalovirus-infected cells

V. Sullivan*, C. L. Talarico†, S. C. Stanat†, M. Davis†, D. M. Coen* & K. K. Biron†

* Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA
† Division of Virology, Burroughs Wellcome Company, Research Triangle Park, North Carolina 27709, USA

HUMAN cytomegalovirus (HCMV) is a major pathogen in immunosuppressed individuals, including patients with acquired immune deficiency syndrome. The nucleoside analogue ganciclovir (9-(1,3-dihydroxy-2-propoxymethyl)-guanine) is one of the few drugs available to treat HCMV infections, but resistant virus is a growing problem in the clinic[1] and there is a critical need for new drugs. The study of ganciclovir-resistant mutants has indicated that the selective action of ganciclovir depends largely on virus-controlled phosphorylation in HCMV-infected cells[2–5]. The enzyme(s) responsible have not been identified. Here we report that the HCMV gene *UL97*, whose predicted product shares regions of homology with protein kinases, guanylyl cyclase and bacterial phosphotransferases[6–8], controls phosphorylation of ganciclovir in HCMV-infected cells. A four-amino-acid deletion of UL97 in a conserved region, which in cyclic AMP-dependent protein kinase participates in substrate recognition[9], causes impaired ganciclovir phosphorylation. The implications of these results for antiviral drug development and drug resistance are discussed.

To map the HCMV-encoded function that controls ganciclovir phosphorylation, the viral genome of mutant 759'D100, which is deficient in drug anabolism[4], was cloned into an overlapping set of nine cosmids, which were then used in marker transfer experiments (Table 1). Two non-overlapping cosmids, pC7S95 and pC7S37 (Fig. 1), transferred ganciclovir resistance to the wild-type parental strain AD169. The resistance marker of pC7S37 has been mapped to the viral DNA polymerase and shown to be unrelated to ganciclovir phosphorylation (V.S. *et al.*, manuscript submitted). In contrast, the ganciclovir-resistant recombinant virus GDG'K17, isolated from marker transfer experiments with pC7S95, gave a value for the 50% effective dose of ganciclovir which was 3-10-fold higher than that of wild-type AD169 (data not shown) and was unable to induce the phosphorylation of ^{14}C-labelled ganciclovir in infected cells (Fig. 2*a*). Similarly, GDG'XBAF4, a ganciclovir-resistant recombinant virus isolated from marker transfer experiments with the F fragment of 759'D100, produced by digestion with the restriction endonuclease *Xba*I (data not shown), was also unable to phosphorylate the drug (Fig. 2*a*). Thus, both a ganciclovir-resistance marker of 759'D100, and the mutation responsible for the drug-phosphorylation defect are contained in the 13-kilobase (kb) overlap of the F fragment and pC7S95 (Fig. 1). Further marker transfer experiments with plasmids containing cloned 759'D100 DNA fragments within the 13-kb overlap localized the ganciclovir-resistant marker to a 2.6-kb region of DNA containing the complete *UL97* and parts of the *UL96* and *UL98* open reading frames (Fig. 1). Ganciclovir anabolism studies of the ganciclovir-resistance recombinant viruses, GDG'SAL4, GDG'HS1 and GDG'EH9, isolated from marker transfer experiments with plasmids pSAL7, pHS7 and pGEH7, respectively (Fig. 2*b*), confirmed that the mutation affecting ganciclovir phosphorylation was also contained in the 2.6-kb fragment. DNA sequence analysis of this region revealed only a single change when compared with the wild-type AD169 sequence: a 12-base-pair (bp) deletion within the *UL97* open reading frame (Fig. 3). These results show that the gene product of *UL97* controls ganciclovir phosphorylation in HCMV-infected cells and that mutations in this open reading frame can confer resistance to this drug.

Homologues of *UL97* are encoded by herpes simplex virus[10], varicella zoster virus[11], Epstein-Barr virus[12] and human herpesvirus 6 (HHV-6), which suggests that there may be a conservation of function among the α, β and γ human herpesviruses[6]. These genes encode regions of homology conserved among protein kinases and phosphotransferases[6,13,14]. The 12-bp deletion in

TABLE 1 Marker transfer of ganciclovir resistance

Transfected DNA	Plating efficiency (%) in ganciclovir
—	<0.04
pC7S2	<0.03
pC7S6	<0.09
pC7S11	<0.17
pC7S31	<0.04
pC7S37	**2.3**
pC7S40	<0.05
pC7S47	<0.23
pC7S95	**2.8**
pC7Sdl	<0.23

Marker transfer of ganciclovir resistance. Human foreskin fibroblasts were transfected with infectious wild-type AD169 DNA alone (—), or cotransfected with AD169 DNA and the indicated cosmid. Progeny virus from the transfected cells were tested for their ability to form plaques in 35 μM ganciclovir (per cent plating efficiency). Marker transfer of ganciclovir resistance by pC7S37 and pC7S95 is shown in bold. The preparation and maintenance of fibroblasts, preparation of infectious AD169 DNA, and details of marker transfer experiments will be described elsewhere (V.S. *et al.*, manuscript submitted). 759'D100 viral DNA was prepared essentially as described for AD169, except MRC-5 cells were used and only extracellular virus was collected. After partial digestion with *Sau*3A, fragments of 20–30 kb were ligated into the *Bam*HI site of cosmid vector pC7108 (ref. 25), packaged into phage (BRL λ packaging system), and cosmid clones isolated by phage infection of *E. coli* (strain N4956) and selected with ampicillin. Cosmids pC7S2, 6, 11, 31, 37, 40, 47, 95 and dl were isolated, characterized by digestion with restriction enzymes and Southern blot analysis (results not shown), and partially digested with *Hin*dIII before use in marker transfer experiments.

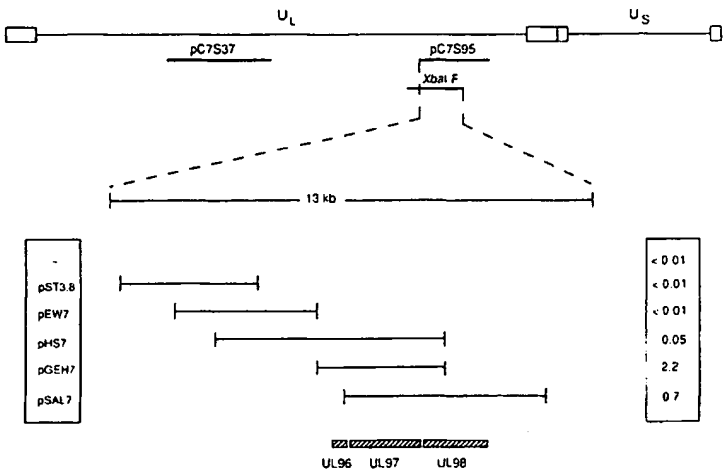

FIG. 1 Molecular mapping of ganciclovir-resistance markers of HCMV. The top line represents the prototype 240-kb HCMV AD169 genome[18]. Boxes represent inverted repeats and the solid line represents unique sequences (U_L, U_S). The solid lines below indicate the genomic position of the 759'D100 *Xba*1 F fragment and cosmids pC7S37, pC7S95; the limits of the 13-kb overlap of pC7S95 and *Xba*1 F are also indicated. The box on the left lists the plasmids used in the marker transfer experiments. The lines in the middle represent the 759'D100 DNA sequences within the 13-kb region contained in the indicated plasmids. The box on the right shows the plating efficiency (%) in 50 μM ganciclovir of progeny virus from human foreskin fibroblasts transfected with AD169 DNA alone (-) or cotransfected with AD169 DNA and each of the plasmids indicated. Each of the plasmids contains DNA fragments subcloned from cosmid pC7S95. pHS7 and pEW7 contain the *Hin*dIII S and *Eco*RI W fragments[26] respectively, cloned into the *Hin*dIII or *Eco*RI sites of pUC18. pST3.8 contains a 3.8-kb *Pst*I fragment inserted into the *Pst*I site of pGEM5Zf(+) (Promega Inc.). pGEH7 contains a 3.5-kb *Eco*RI/*Hin*dIII fragment inserted into the *Eco*RI and *Hin*dIII sites of pGEM7Zf(+). pSAL7 contains a 5.6-kb *Sal*I fragment ligated into the *Xho*I site of pGEM7Zf(+). The 759'D100 DNA fragments were excised from the plasmid vectors with the appropriate restriction enzymes before use in marker transfer experiments. Shaded box represents the 2.6-kb DNA overlap (extending from the *Sal*I to the *Hin*dIII sites) of plasmids pHS7, pGEH7 and pSAL7 which were able to transfer ganciclovir resistance. Hatched boxes below indicate the positions of the *UL96*, *UL97* and *UL98* open reading frames.

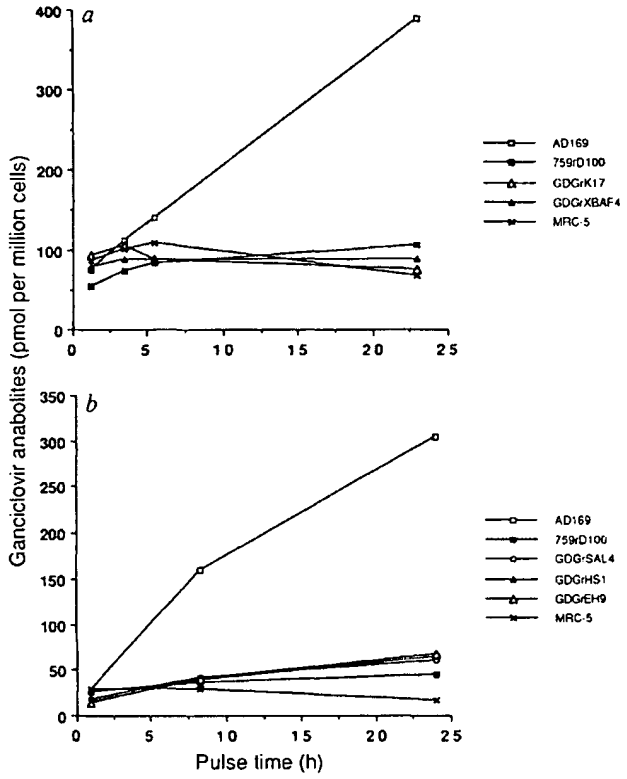

FIG. 2 Phosphorylation of ^{14}C-labelled ganciclovir in HCMV-infected cells. *a*, MRC-5 cells (human embryonic lung fibroblasts), maintained as described[2], were infected with AD169 (□), 759'D100 (■), GDG'K17 (△), or GDG'XBAF4 (▲) at a multiplicity of infection (m.o.i.) of 0.1–0.5, or mock-infected (×). Four days post-infection they were pulse-labelled for 1.25, 3.5, 5.5 and 23 h with 50 μM ^{14}C-labelled ganciclovir (specific activity, 52mCi per mmol) and, after extraction with perchloric acid, the ganciclovir anabolites were measured with a cation-exchange column as described[27]. *b*, MRC-5 cells were infected with AD169 (□), 759'D100 (■), or GDG'SAL4 (○), GDG'HS1 (▲), GDG'EH9 (△), at m.o.i. 0.1–0.5, or mock infected (×). Four days post-infection they were labelled with ^{14}C-labelled ganciclovir for 1, 8.3 and 24 h, and the levels of ganciclovir anabolites determined.

UL97 results in a 4-amino-acid deletion in the UL97 product of residues 638–641 (Ala-Ala-Cys-Arg) (Fig. 3). These residues lie in a region conserved among protein kinases, designated subdomain IX (ref. 14) or region VI (ref. 13). The conservation of this region indicates that it may have an important and common function in these enzymes[13,14]. X-ray crystallography of the catalytic subunit of bovine cAMP-dependent protein kinase has shown that residues corresponding to those deleted in UL97 form a hydrophobic binding site for a peptide inhibitor (PKI (5–24)) and, by extension, its substrate[9]. The deletion of residues at positions implicated in substrate binding of protein kinases raises the possibility that the mutant 759'D100 *UL97* product may retain enzyme activity with altered substrate specificity. It is interesting to note that three of the deleted amino acids are conserved in the HHV-6 homologue of UL97 (15R in Fig. 3). HHV-6 is also sensitive to ganciclovir[15,16] and, like HCMV, no homologue of the thymidine kinase responsible for ganciclovir phosphorylation in cells infected by herpes simplex virus[17] has been found in this virus[15,16,18]. This raises the possibility that the HHV-6 homologue of UL97 may be responsible for ganciclovir phosphorylation in cells infected by HHV-6.

The role of UL97 in the HCMV replication cycle is not known. The predicted product is obviously related to protein kinases and kinase activity has been reported in HCMV-infected cells and virions[19–23], but no protein kinase function has yet been associated with UL97.

On the basis of these results, we envisage two possible mechanisms by which ganciclovir phosphorylation is controlled by UL97. UL97 may phosphorylate another enzyme, thereby activating this second enzyme to phosphorylate ganciclovir; although we cannot exclude this model, we favour a different mechanism in which UL97 itself phosphorylates the drug. This simpler mechanism is plausible because UL97 departs from consensus protein kinase motifs at the same residues that also vary in bacterial aminoglycoside phosphotransferases[7] and in sea urchin integral membrane guanylyl cyclase[8], enzymes that transfer phosphates to sugar hydroxyl moieties. Ganciclovir is phosphorylated on its acyclic sugar-like moiety. Indeed, Chee *et al.*, who pointed out these divergences, have suggested that UL97 deserves consideration as the HCMV-encoded ganciclovir kinase[6], a proposal supported by evidence in the accompanying paper[24] showing that UL97 can carry out this enzymatic function.

The mapping of an HCMV ganciclovir-resistance marker to *UL97* identifies a new selective target which should facilitate the development of safer and more effective anti-HCMV drugs.

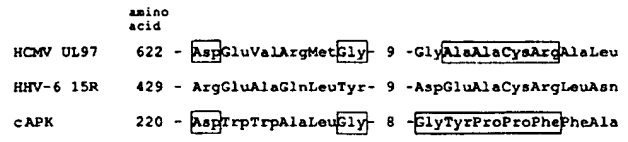

FIG. 3 A four-amino-acid deletion in the HCMV *UL97* gene product confers resistance to ganciclovir. Subdomain IX of bovine cAMP-dependent protein kinase (cAPK) (amino acids 220–240) (refs 14, 28) and the corresponding amino acids in HCMV UL97 (amino acids 622–643) and HHV-6 15R (amino acids 429–450) are shown[6]. The four amino acids deleted in ganciclovir-resistant mutant 759'D100 are darkly shaded. The five residues in subdomain IX of cAPK involved in substrate binding are lightly shaded[9]. Residues highly conserved in protein kinases are outlined[14]. The DNA sequence of the 2.6-kb *Sall*/*Hin*dIII fragment of 759'D100 was obtained by sequencing plasmid pHS7 as double-stranded DNA using a commercial kit (Sequenase version 2.0, US Biochem.) and oligonucleotide primers (5'-TGAGTCGTCGTTC-CGA-3',5'-TTGTGCTGACGGAACTCA-3',5'-AACGGGCGTTTGCGTCCA-3',5'-AGC-AGGTAGGAGATAGAG-3',5'-CTTCCACTGGTCTGTAT-3',5'-ATTCGTGCAGCATG-GTCT-3' (VS976), 5'-TACGGCGTTATTGCATGT-3' (VS977), 5'GGTAACATTCG-CGCAGA-3',5'-ACACCCATGAACGTGCT-3',5'-CTTGGTGGACTCGGTTTC-3',5'-GAAACTTCGGCCATGTC-3',5'-GTTTGTACCTTCTCTGTT-3') (Oligos Etc.) designed using the published AD169 sequence[6]. A 12-bp deletion 5'-GCGGCC-TGCCGC-3' was found in the *UL97* coding region in pHS7 and was also shown to be deleted in plasmids pGEH7 and pSAL7. Direct sequencing of this region, after amplification by polymerase chain reaction using oligonucleotide primers VS976 and VS977, confirmed the presence of the 12 bp in the AD169 infectious DNA used in the marker transfer experiments and their absence in 759'D100, GDG'HS1, EH9 and SAL4 viral DNAs. One other change compared with the published AD169 sequence was found: only nine, instead of the published ten A residues, were found in the non-coding region 69 bp upstream of *UL97*, but this was not considered significant as sequence analysis of cloned[26] or infectious AD169 DNA in this region also contained only nine A residues.

Resistance of HCMV to ganciclovir in the clinic is predominantly due to impaired drug phosphorylation[5], presumably as a result of UL97 mutation. It is important to determine the role of UL97 in the pathogenicity of the virus. If UL97 proves to be essential to the virus, then ganciclovir or a new drug might act not only as a substrate for this enzyme but also as an inhibitor of its normal function in the virus replication cycle. Independently, UL97 may help define residues and structures important for catalytic specificity during the evolution of a protein kinase homologue into an enzyme with a different function. □

Received 11 February; accepted 26 May 1992.

1. Drew, W. L. *et al. J. infect. Dis.* **163**, 716-719 (1991).
2. Biron, K. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2473-2477 (1985).
3. Freitas, V. R., Smee, D. F., Chernow, M., Boehme, R. & Matthews, T. R. *Antimicrob. Ag. Chemother.* **28**, 240-245 (1985).
4. Biron, K. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 8769-8773 (1986).
5. Stanat, S. C. *et al. Antimicrob. Ag. Chemother.* **35**, 2191-2197 (1991).
6. Chee, M. S., Lawrence, G. L. & Barrell, B. G. *J. gen. Virol.* **70**, 1151-1160 (1989).
7. Brenner, S. *Nature* **329**, 21 (1987).
8. Singh, S. *et al. Nature* **334**, 708-712 (1988).
9. Knighton, D. R. *et al. Science* **253**, 414-420 (1991).
10. McGeoch, D. J. *et al. J. gen. Virol.* **69**, 1531-1574 (1988).
11. Davison, A. J. & Scott, J. E. *J. gen. Virol.* **67**, 1759-1816 (1986).
12. Baer, R. *et al. Nature* **310**, 207-211 (1984).
13. Smith, R. F. & Smith, T. F. *J. Virol.* **63**, 450-455 (1989).
14. Hanks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42-52 (1988).
15. Agut, H. *et al. Res. Virol.* **140**, 219-228 (1989).
16. Di Luca, D., Kalsafanas, G., Schrimer, E. C., Balachandran, N. & Frenkel, N. *Virology* **175**, 199-210 (1990).
17. Fyfe, J. A., Keller, P. M., Furman, P. A., Miller, R. L. & Elion, G. B. *J. biol. Chem.* **253**, 8721-8727 (1978).
18. Chee, M. S. *et al. Curr. Top. Microbiol. Immun.* **154**, 125-169 (1990).
19. Mar, E.-C., Patel, P. C. & Huang, E.-S. *J. gen. Virol.* **57**, 149-156 (1981).
20. Michelson, S., Tardy-Panit, M. & Barzo, O. *Virology* **134**, 259-268 (1984).
21. Roby, C. & Gibson, W. *J. Virol.* **59**, 714-727 (1986).
22. Britt, W. J. & Auger, D. *J. Virol.* **59**, 185-188 (1986).
23. Somogyi, T., Michelson, S. & Masse, M.-J., O. *Virology* **174**, 276-285 (1990).
24. Littler, E., Stuart, A. D. & Chee, M. S. *Nature* **358**, 160-162 (1992).
25. Lau, Y.-F. & Kan, Y. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5225-5229 (1983).
26. Fleckenstein, B., Muller, I. & Collins, J. *Gene* **18**, 39-46 (1982).
27. Fyfe, J. A., McKee, S. A. & Keller, P. M. *Molec. Pharmac.* **24**, 316-323 (1983).
28. Shoji, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 848-851 (1981).

ACKNOWLEDGEMENTS. We thank K. Easley and H. Green for help in obtaining human foreskins; S. Weinheimer for infectious AD169 DNA; K. Ruffner, E. Bodin, Y. Mehrotra, L. F. Mattingly and S. A. Short for technical assistance; S. Brenner for discussion; E. Littler, A. D. Stuart and M. S. Chee for making unpublished data available; and the NIH for grant support.

Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases

W. Bode, F. X. Gomis-Rüth, R. Huber, R. Zwilling* & W. Stöcker*

Max-Planck-Institut für Biochemie, W-8033 Martinsried, Germany
* Zoologisches Institut, Universität Heidelberg, Im Neuenheimer Feld 230, W-6900 Heidelberg, Germany

ASTACIN, a digestive zinc-endopeptidase from the crayfish *Astacus astacus* L.[1,2], is the prototype for the 'astacin family'[3–5], which includes mammalian metallo-endopeptidases[5] and developmentally regulated proteins of man[6], fruitfly[7], frog[8] and sea urchin[9,10]. Here we report the X-ray crystal structure of astacin, which reveals a deep active-site cleft, with the zinc at its bottom ligated by three histidines, a water molecule and a more remote tyrosine. The third histidine (His 102) forms part of a consensus sequence, shared not only by the members of the astacin family, but also by otherwise sequentially unrelated proteinases, such as vertebrate collagenases[11]. It may therefore represent the elusive 'third' zinc ligand in these enzymes. The amino terminus of astacin is buried forming an internal salt-bridge with Glu 103, adjacent to His 102. Astacin pro-forms extended at the N terminus, as observed for some 'latent' mammalian astacin homologues, did not exhibit this 'active' conformation, indicating an activation mechanism reminiscent of trypsin-like serine proteinases.

We have solved the X-ray crystal structure of astacin by multiple isomorphous heavy-atom replacement and have refined its structure to a crystallographic *R*-value of 0.162 using reflections from 10 to 1.8 Å resolution. The structure reveals that astacin is a spherical molecule, subdivided by a long, deep active-site cleft into two domains (Fig. 1*a*). The 'N' domain (Fig. 1*b*) comprises the N-terminal chain up to residue 100 and consists mainly of a five-stranded pleated β-sheet and two long α-helices. The carboxy-terminal 'C' domain is largely organized in several turns and irregular structures (Fig. 1*b*). The N terminus is buried in the lower domain, whereas the C terminus extends towards the 'N' domain, covalently linked to the latter by a disulphide bridge between Cys 42 and Cys 198. The catalytic zinc resides at the bottom of the cleft, liganded by His 92, His 96 and His 102, a water molecule (Sol 300) anchored to Glu 93, and the OH of Tyr 149 (Fig. 2) in a trigonal-bipyramidal manner. His 102 Nε2 and Tyr 149 OH, at distances of 2.3 and 2.6 Å to the zinc, respectively, form the vertices, whereas the other three ligands are coplanar with, and 2.1 Å apart from, the central zinc. His 92, Glu 93 and His 96 are located on a central α-helix, which extends up to Gly 99 where the chain turns sharply towards His 102 (Fig. 2).

The side-chain of Glu 103, adjacent to the liganding His 102, points away from the active-site zinc into a water-filled cavity inside the 'C' domain (Fig. 3). Its carboxylate group is packed against some internal solvent molecules; one of them (Sol 501) mediates through hydrogen-bonds a salt-bridge to the ammonium group of the N-terminal Ala 1, which is completely buried in the interior of the molecule (Fig. 3).

Until now, the only available crystal structures for zinc-endopeptidases have been those for thermolysin[12] and two closely related bacterial enzymes[13,14]. In these enzymes, two histidine residues, a glutamic acid and a (glutamic acid-bound) water molecule are ligands to the active-site zinc ion. Although of a different topology, the same arrangement of ligands are seen in the exopeptidase carboxypeptidase A (ref. 15).

The five-stranded β-sheet and both helices of the 'N' domain of astacin are topologically similar to thermolysin, although the